Metabolomics (2009) 5:239–252
DOI 10.1007/s11306-008-0146-y

ORIGINAL ARTICLE

Metabolic profiling of transgenic wheat over-expressing the high-molecular-weight Dx5 glutenin subunit

Boryana S. Stamova · Ute Roessner · Suganthi Suren · Debbie Laudencia-Chingcuanco · Antony Bacic · Diane M. Beckles

Received: 12 April 2008 / Accepted: 13 November 2008 / Published online: 9 December 2008

Abstract The primary aim of this work was to evaluate potential changes in the metabolic network of transgenic wheat grain over-expressing the high-molecular-weight (HMW) glutenin Dx5-subunit gene. GC–MS and multivariate analyses were used to compare the metabolite profiles of developing caryopses of two independently transformed lines over-expressing *Dx5* and another two independently transformed lines expressing only the selectable-marker gene (controls). Developing grain at 7, 14 and 21 Days Post-Anthesis (DPA) was studied to observe differences in metabolically active tissues. There was no distinction between the *Dx5* transformants and the controls by principal component analysis (PCA) suggesting that their metabolite compositions were similar. Most changes in metabolite levels and starch occurred at 14 DPA but tapered off by 21 DPA. Only 3 metabolites, guanine, 4-hydroxycinnamic acid and Unknown 071306a, were altered due to *Dx5* expression after correction for false discovery rates ($P < 0.0005$). However, discriminant function analysis (DFA) and correlative analyses of the metabolites showed that Dx5-J, which had the highest level of Dx5 protein in ripe caryopses, could be distinguished from the other genotypes. The second aim of this work was to determine the influence of gene transformation on the metabolome. Cross-comparison of the transformed controls to each other, and to the *Dx5* genotypes showed that approximately 50% of the metabolic changes in the *Dx5* genotypes were potentially due to variations arising from gene transformation and not from the expression of the Dx5-gene per se. This study therefore suggests the extent to which plant transformation by biolistics can potentially influence phenotype.

Keywords Transgenic wheat · Storage protein · GC–MS · Multivariate analysis

Electronic supplementary material The online version of this article (doi:10.1007/s11306-008-0146-y) contains supplementary material, which is available to authorized users.

B. S. Stamova · D. M. Beckles (✉)
Department of Plant Sciences-Mail Stop 3,
University of California-Davis, 1 Peter Shields Avenue,
Davis, CA 95616, USA
e-mail: dmbeckles@ucdavis.edu

B. S. Stamova · D. Laudencia-Chingcuanco
Genomics and Gene Discovery Unit, USDA-Albany,
800 Buchanan Street, Albany, CA 94710, USA

B. S. Stamova
Genetics Resources Conservation Program, University of
California-Davis, 1 Shields Avenue, Davis, CA 95618, USA

Present Address:
B. S. Stamova
M.I.N.D Institute, Department of Neurology, School
of Medicine, University of California Medical Center,
2805 50th Street, Sacramento, CA 95817, USA

U. Roessner · S. Suren · A. Bacic
Australian Centre for Plant Functional Genomics and
Metabolomics Australia, School of Botany, University
of Melbourne, Parkville 3010, VIC, Australia

1 Introduction

Wheat is one of the most important crops cultivated globally (FAO 2006). It is valued for the gluten in the grain which enables its use for bread and pasta-making (Vasil 2007). An important component of wheat gluten is the Dx5 protein, a high-molecular-weight glutenin subunit (HMW-GS) (Shewry et al. 2002). There is a good correlation between dough visco-elasticity and the amount of the

HMW-GS in wheat endosperm (Payne et al. 1987; Shewry et al. 2002, 2003). Accordingly, the HMW-GS genes were some of the first to be manipulated by gene transformation in order to define their role in determining dough characteristics and to possibly engineer wheat with enhanced functionality in food (Blechl et al. 1998; He et al. 1999; Alvarez et al. 2000; Barro et al. 2002; Altpeter et al. 2004).

Plant gene transformation is done usually by biolistics or *Agrobacterium*-mediated transfer and both can introduce unintended genetic changes to the organism. Multiple and randomly located gene insertions can occur, and the introduced plasmid vector can become re-arranged or truncated (Barcelo et al. 2001; Rooke et al. 2003). Biolistics, the most frequently used method for wheat transformation thus far (Fu et al. 2007), can produce as many as 40 copies of the construct in a single site, as well as multiple copies in random places in the genome (Vain et al. 2002). These events have the potential to influence phenotype especially if they disrupt the function of a gene(s) by insertion. Furthermore, some characteristics in the transformed progeny are due to the epigenetic and genetic changes that can occur during tissue culture as well as the expression of the selectable marker often engineered into the plasmid vector (Filipecki and Malepszy 2006). Therefore, independent of the gene sequence introduced, lines produced by gene transformation may manifest phenotypes that are due to a combination of several factors including changes in plasmid architecture, the site(s) of plasmid integration, selectable marker expression as well as somaclonal variation.

Molecular large-scale profiling technologies, i.e. transcriptomics, proteomics and metabolomics are useful tools for evaluating compositional changes in transgenic crops due to gene transformation (Baudo et al. 2006). Metabolites, as the final products of gene activity, are good indicators of meaningful biochemical changes that result from transcriptional re-programming (Sweetlove and Fernie 2005). Metabolite profiling was used by Baker et al. (2006) to determine the effect of altering the HMW-GS Dx5 subunits and other wheat gluten proteins on the biochemical composition of ripened grain (Baker et al. 2006). They used NMR and GC–MS to profile metabolites in ripened grain sampled from plants grown over different years and in various locations. The different transgenic and control lines in that study could not be distinguished by principal component analysis (PCA) suggesting that there were few changes to the metabolome resulting from manipulation of storage proteins composition in the ripe caryopsis.

We are interested in understanding how changing storage protein biosynthesis in wheat caryopses might affect the primary biochemical pathways in that tissue. To address this, we assayed the polar components of the developing grain in two transgenic wheat lines with additional copies of the *HMW-Dx5* gene and in two control lines containing only the selectable-marker gene, *bar*, by GC–MS. The wheat lines we studied were previously analysed for changes in agronomic traits (Bregitzer et al. 2006), protein levels and dough functional properties (Blechl et al. 2007). Over-expression of *Dx5* in these lines did not alter yield or total seed protein; however, there was a redistribution of the individual components of the gluten fraction with a change in dough properties (Blechl et al. 2007). When taken together, the independent studies of differently produced transgenic lines over-expressing *Dx5* by (Baker et al. 2006) and (Blechl et al. 2007) showed that there were only minor changes due to the modification and that physiological adjustments might have been limited to storage protein biosynthesis with potentially minimal impact on other pathways.

The primary aim of the work described in this paper therefore, was to determine specifically, the way in which metabolism was adjusted in order to compensate for over-expression of *Dx5* in developing transgenic caryopses. We analysed the data by PCA, Independent Component Analysis and Student's *t*-test to compare the genotypes and determine if they are metabolically similar as indicated by past research (Baker et al. 2006). Developing grain at 7, 14 and 21 DPA were studied to determine if development influences the response to the perturbation. Baker et al. (2006) used fully mature grain in their work and so examining developing tissue may yield new information. We also used additional multivariate analyses not performed by Baker et al. (2006) in an attempt to see if the *Dx5* transformants and the controls could be differentiated from each other. Correlative matrices of metabolites were drawn for each genotype in order to build up a picture of their metabolic network and to identify how they were altered by over-expression of *Dx5*. A secondary aim of this work was to determine the extent to which genetic and epigenetic changes associated with gene transformation influenced the phenotype of the transgenic lines. For these analyses, the two independently transformed control wheat lines were compared to each other and then to the *Dx5* transformants, which gave some insight on metabolite changes not associated with *Dx5* over-expression.

2 Materials and methods

2.1 Plant material

Transformation of wheat lines with the gene encoding the HMW-Dx5 subunit was by particle bombardment as described previously (Bregitzer et al. 2006; Blechl et al. 2007). The lines BAR-C and BAR-D were transformed independently with a single plasmid-type containing the

selectable-marker gene, *bar*, under the control of the ubiquitin promoter (Christensen and Quail 1996) (Table 1). The Dx5-G and Dx5-J lines were independently transformed with a mixture of the two types of plasmids; one plasmid contained the native *Dx5* gene under the control of its own promoter and the other plasmid contained the same chimeric selectable-marker gene (*ubi-bar*) as in BAR-C and BAR-D. Dx5-G had 2.3-fold and Dx5-J had 3.5-fold the level of Dx5 protein in the mature caryopses than a non-transformed control, while the Dx5 levels in BAR-C and BAR-D genotypes were similar to the same controls (Blechl et al. 2007). Each line was grown under greenhouse conditions as outlined in (Laudencia-Chingcuanco et al. 2007). All tissue was sampled from the T_6 generation. Whole caryopses, which include the pericarp, embryo and endosperm, were harvested at 7, 14 and 21 DPA and immediately flash-frozen in liquid nitrogen and stored at −80°C until analyzed.

2.2 GC–MS and starch analysis

GC–MS analysis was done on a total of 60 samples. Two individual caryopses were harvested from 5 plants, for each genotype at the three developmental stages. Samples were homogenised using a mortar and pestle, pre-cooled with liquid nitrogen and the powder was extracted in 500 µl of methanol; 20 µl of polar internal standard (0.2 mg ml^{-1} ribitol and norleucine in water) was added as a quantification standard. The mixture was extracted for 15 min at 70°C and subsequently mixed vigorously with 1 volume of water. After centrifugation at 2200*g*, the supernatant was transferred into a new tubes and 2 aliquots, each of 100 µl, were taken and dried in vacuo for further derivatisation with TBS or TMS (Jacobs et al. 2007). Sample derivatisation and GC–MS analyses were carried out as previously described (Roessner et al. 2001, 2006; Jacobs et al. 2007). Raw data was expressed as the relative response ratios per gram fresh weight with the area of each analyte normalised to the area of the internal standard, ribitol (for TMS-derivatised samples) or norleucine (TBS-derivatised samples), and the fresh weight of each sample. Starch was assayed on caryopses sampled from the same plants used for metabolite profiling as described (Beckles et al. 2001).

Table 1 Plasmid constructs used in this study

Line name	Plasmid 1 Promoter:Transgene	Plasmid 2 Promoter:Transgene	Dx5 protein level
BAR-C	–	*Ubi*:*bar*	x1
BAR-D	–	*Ubi*:*bar*	x1
Dx5-G	*Dx5*:*Dx5*	*Ubi*:*bar*	x2.3
Dx5-J	*Dx5*:*Dx5*	*Ubi*:*bar*	x3.5

The genotypes used in this study were the T_6 generation of lines produced by microprojectile bombardment with either one (genotype BAR-C and BAR-D) or two (Dx5-G and Dx5-J) plasmids. The selectable marker was the *bar* gene which encodes resistance to the herbicide bialaphos under the control of the ubiquitin promoter (Christensen and Quail 1996). These genotypes were chosen on the basis that they showed little variation in morphology or growth when compared to the non-transformed controls. For a complete description of plasmids and transformation events, as well as data on Dx5 protein levels, please refer to previous publications (Blechl and Anderson 1996; Bregitzer et al. 2006)

2.3 Statistical and multivariate analysis of the dataset

The data matrix was calculated using the formula $M \times N$ where M is the number of metabolites measured (109) and N is the number of samples in the experiment (5 biological replicates × 4 genotypes × 3 developmental stages). In this experiment the entire dataset was $M \times N = 109 \times 60 = 6540$. Statistically significant differences in metabolite levels between samples were identified using the Student's *t*-test at $P < 0.05$ level (Steel et al. 1997). When making parallel comparisons between many samples, the possibility of falsely rejecting the null hypothesis, i.e. determining that there is no statistically significant difference between samples increases significantly (Broadhurst and Kell 2006). The Bonferroni correction was applied to the generated data to detect those metabolite levels which differed after correcting for false discovery (Abdi 2007). The Bonferroni correction is a more conservative approach to reducing false discovery. The *P*-value is lowered from 0.05 to 0.05/109 (number of metabolites measured) or $P < 0.0005$ (Abdi 2007).

Pearson's product-moment correlation coefficients were calculated from the data matrix using the response ratios for each metabolite (Morgenthal et al. 2006). These correlations are given as *r*-values where "1" indicates a perfect correlation, "0" no correlation and "−1" shows an inverse correlation between variables. The number of metabolite-to-metabolite pairs analysed for the correlative matrices of each genotype was $[(109*109)/2 - (109/2)] = 5886$. Correlative analyses were computed using Microsoft Excel (Seattle, WA), heat maps were displayed using JCOLORGRID (Joachimiak et al. 2006).

Multivariate analyses were done after the data was $\log_{10}$-transformed in order to bring the values closer to a Gaussian distribution. Principal component analysis was done using Statistical Analysis Software (SAS Inc, Cary NC). Discriminant function analysis (DFA) and Independent Component Analysis (ICA) were done using Statistica Data Miner Software (Statsoft Inc., Tulsa, OK; Statsoft 2003). DFA is a supervised statistical algorithm that will derive an optimal separation between groups established a priori using within-group variances and covariance

matrices developed from each group of data (Raamsdonk et al. 2001). The a priori groups were defined as the different wheat genotypes in this study. The significance level of the discriminant functions were determined using the Wilks' Lambda (Statsoft 2003). For ICA, the dimensionality of the data was first reduced using PCA (Scholz et al. 2004). The first nine PCs, which described 80% of the variance between samples, were extracted and then analysed using fast ICA (Scholz et al. 2004).

3 Results and discussion

3.1 Principal component analysis (PCA) of metabolites

The primary aim of this work was to determine the extent to which the four transgenic wheat genotypes studied were similar to each other based on the polar metabolite composition of their caryopses. Lines BAR-C and BAR-D were transformed only with the *bar*-plasmid while Dx5-G and Dx5-J were co-bombarded with a mixture of the *ubi-bar* and the *Dx5*-plasmid (Table 1). Caryopses were sampled from five individual plants of each genotype at 7, 14 and 21DPA, and methanol extracts were analysed by GC–MS. The entire dataset consisted of 60 samples. The relatedness of the four transgenic wheat genotypes to each other was first examined using PCA. PCA is an unsupervised multivariate method that allows patterns, trends, groups and outliers in large datasets to be easily identified. The dimensionality of complex data is reduced to what are called Principal Components (PC) that retain the maximal amount of variation within a sample. The first PC captures the most variation; the second PC captures the next level of variation and so on. For a detailed description of PCA the reader is referred to Joliffe (1986). When the entire dataset was analysed, there was no distinction between the *Dx5*- and *bar*-only-transformed genotypes (Fig. 1). The 60 samples fell into classes based on caryopsis age (Fig. 1). There was also a high level of plant-to-plant variation evident from the PCA pattern. No separation of the genotypes was observed when PCA was applied to the data either at each developmental stage separately (Supplemental Fig. 1a–c), or using the means of the five individual biological replicates for each genotype (data not shown).

The contribution of individual metabolites to the PCA output in Fig. 1 was estimated from the PCA loading score plot (Fig. 2). Metabolites that cluster around the origin of a PCA loading plot make little contribution to the PCA separation, whereas outlying metabolites have a greater impact. Many of the metabolites grouped together (Fig. 2), but away from the origin, which implies that all of these metabolites contributed to the separation seen in Fig. 1. Within this cluster are 12 metabolites with the highest

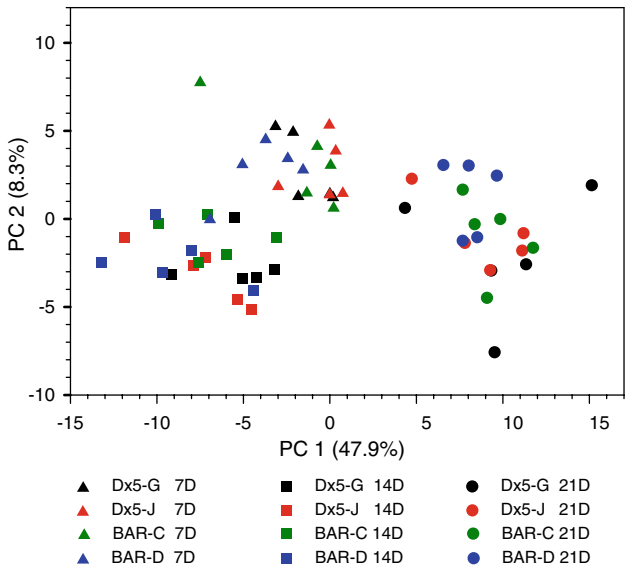


Fig. 1 Principal component analysis (PCA) of 109 polar metabolites measured by GC–MS, in wheat caryopses harvested from the Dx5-G, Dx5-J, BAR-C and BAR-D genotypes at 7, 14 and 21 DPA. Each of the five biological replicates was individually plotted and the samples were projected onto bi-plots showing the first two principal components (PCs). Each symbol on the plot represents data from 109 metabolites reduced to a single data point defined by the first (PC1) and second (PC2) principal components. Samples that have similar metabolite composition will cluster together while samples that are different will be further apart. The percentage variation explained by each individual PC is shown in parentheses on each axis

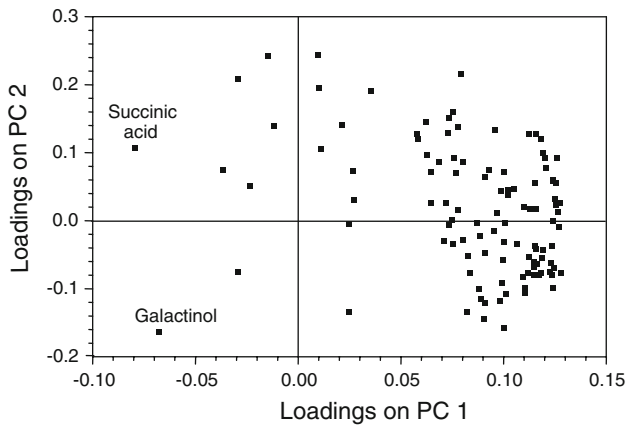
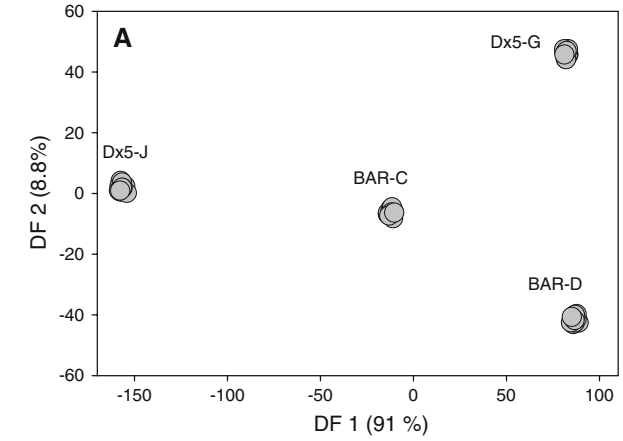


Fig. 2 PCA loading scores of 109 metabolites from the transgenic wheat genotypes Dx5-G, Dx5-J, BAR-C and BAR-D. The loading scores depicted were derived from the PCA output shown in Fig. 1. The loading score of each variable (metabolite) defines the extent to which it participates in determining the separation seen in a PC plot. Each point on the graph shows a metabolite and the contribution of that metabolite to the PCA output. Metabolites that cluster close to the origin (zero) make no contribution, while outliers are presumed to make a greater contribution to the classification of the data observed

loading scores ranging from 0.125 to 0.128 on the first PC (Fig. 2); these included three sugars, two sugar acids and seven fatty acids (Supplemental data Table 1). In contrast,



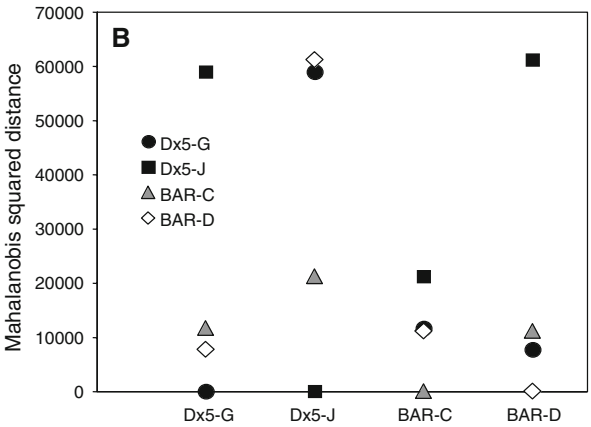
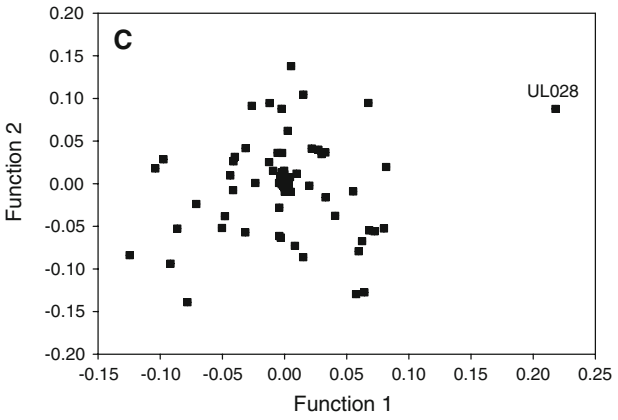


Fig. 3 Discriminant function analysis (DFA) of 109 measured in transgenic wheat caryopses. **a** DFA plots of Dx5-G, Dx5-J, BAR-C and BAR-D genotypes. Samples were projected onto bi-plots of the first two discriminant factors which accounted for 98.8% of the variance observed. **b** Mahalanobis distances between genotypes. The Mahalanobis distance calculates the position of the genotypes from the centroids, or averages of their summed distances, in multidimensional space. Each genotype is set at the baseline in turn and the relatedness of the other genotypes is determined by how far away they are positioned from that baseline. The statistical significance tests of the Class-squared Mahalanobis distances between groups are also shown. The *F*-value indicates the distance between groups and measures the likelihood that the variances are different. The *P*-level is the significance of that distance and estimates the probability that an observed difference between groups occurred by chance alone. **c** Contribution of each metabolite to the separation of genotypes by DFA. Metabolites that contributed most to the separation of the genotypes can be determined by the factor-structure coefficients. The factor-structure coefficients describe the correlation between the metabolites measured and the discriminant function, in this case using the first two discriminant functions. Only one metabolite was an outlier (UL028) and this compound is presumed to have made the greatest contribution to the classification seen in Fig. 4a and b

	Dx5-G	Dx5G	Dx5J	Dx5J	BAR-C	BAR-C	BAR-D	BAR-D
	F	*p*	F	*p*	F	*p*	F	*p*
Dx5-G			141.0	0.067	27.9	0.150	18.5	0.183
Dx5-J	140.9	0.067			50.7	0.111	146.3	0.066
BAR-C	27.8	0.150	50.7	0.111			26.6	0.153
BAR-D	18.5	0.183	146.3	0.066	26.6	0.153		



galactinol and succinic acid were outliers (Fig. 2) indicating that they may serve as developmental biomarkers in these genotypes.

It is possible that PCA was not able to extract all of the possible useful information from the dataset because of the limitation of the technique (Kermit and Tomic 2003; Scholz et al. 2004). We therefore used Independent Component Analysis (ICA), which is often superior to PCA, in an attempt to achieve better separation of the genotypes. The ICA plots did not result in separation of the samples based on genotype (data not shown) and as such did not provide any further insight into the relationships of the lines we examined. We were unable to discriminate between the transgenic lines when applying unsupervised clustering tools to the dataset. Our global analysis of polar metabolite levels is therefore largely in agreement with that performed by Baker et al. (2006), in spite of the differences in the tissues and developmental stages used.

3.2 Discriminant function analysis (DFA) of genotypes

DFA is a supervised statistical algorithm that will derive an optimal separation between groups established a priori by maximizing between-group variance while minimizing within-group variances (Raamsdonk et al. 2001). We defined the a priori groups as the different wheat genotypes. Using DFA, the wheat lines could be clearly classified by genotype with the first discriminant function (DF) accounting for 91% of the variance between samples (Fig. 3a). The statistical significance of the DFs were calculated using the Wilks' Lambda and were found to be highly significant ($P = 0.00000$ for DF1 and DF2, and $P = 0.000455$ for DF3; Supplemental Table 2).

We analysed the potential effect of variance on the DFA. The total within-group variance was higher in the two Dx5-transformed genotypes when compared to the vector-only lines, and was highest in Dx5-J. There was also more variance at 7DPA than at 14 or 21DPA for all the genotypes studied (data not shown). When the DFs were

re-plotted taking into account the variance from developmental stage, Dx5-J still was most disparate to the other genotypes (data not shown). Further the variance found in Dx5-G (12599) was similar to that in Dx5-J (13780), which were both higher than that in BAR-C (11089) and BAR-D (9280). Yet, by DFA, Dx5-G clustered together with the controls and Dx5-J was an outlier. Variance alone is unlikely to be the sole reason for the divergence of the Dx5-J from the other genotypes in this analysis.

To quantify the relationship of the genotypes to each other, the DFs established for each genotype were averaged to obtain what is described as the group "centroids" (Statsoft 2003). The distance from which each group separates from the centroids is an estimate of their relatedness; genotypes that are most similar will cluster together whereas those that are least related will be more distant from each other (Fig. 3b). From this assessment, it can be deduced that BAR-C is similarly related to all of the other genotypes, Dx5-G and BAR-D are more related to each other, whereas Dx5-J is least related to all the other genotypes (Fig. 3b). Tests of the statistical significance of the distances (the Class-squared Mahalanobis distance) between groups are also shown (See Fig. 3b legend). The *P*-level estimates the probability that an observed difference between groups occurred by chance alone. The *P*-value for Dx5-J vs. BAR-D and Dx5-G is 0.066 (only slightly higher than 0.05) and, although not at the statistically significant level, it does support the observation that Dx5-J is different from the other genotypes.

Next, the metabolite(s) responsible for the separation seen in the DFA plot were identified. The factor-structure coefficients describe the correlation between the metabolites and the DF and are similar to PCA loading scores. Plots of these coefficients show that Unknown compound, UL028, was the only outlier of the 109 metabolites measured and, thus, differences in the level of this compound probably underscored the separation of the genotypes by DFA (Fig. 3c).

Overall the DFA results suggest that while the genotypes are similar, Dx5-J which has the highest level of Dx5 protein can be discriminated from the other genotypes. There was a discrepancy in the activity of the *bar*-gene in Dx5-J line; although the presence of the *bar* gene was confirmed by PCR, Dx5-J did not show the same level of resistance to glufosinolate compared to the other genotypes (Ann Blechl; personal communication). The extent to which this influenced the results is not known. Nevertheless, it seems more plausible that the high level of Dx5 in Dx5-J gluten, which caused severe changes in dough viscoelastic properties of Dx5-J flour, resulted in many of the differences among genotypes captured by DFA (Blechl et al. 2007).

3.3 Metabolites levels in the wheat genotypes

Metabolites that differed in absolute content between the *Dx5*-transformed genotypes and the controls were identified using the Student's *t*-test and were painted onto a metabolic map (4A-C). Parametric tests, like the Student's *t*-test, assume that the data have been sampled from a population that approximates to a Gaussian distribution. Non-parametric tests do not have this assumption of normality, however, they are less powerful, particularly with small sample sizes where the *P*-values tend to be higher. This makes it harder to identify real differences as being statistically significant. The data was log-transformed in order to bring the values closer to a Gaussian distribution thus making the use of parametric tests feasible.

The levels of several intermediates of central metabolism were affected as a result of *Dx5* over-expression, including some associated with the TCA cycle, amino acid and fatty acid metabolism (Fig. 4a–c). Perhaps surprisingly, there were comparatively fewer changes associated with carbohydrate metabolism. Differences in metabolite levels among genotypes were modest, with the greatest magnitude generally less than fourfold (data not shown). We checked the likelihood that compounds that differed significantly between genotypes were false positives. The Bonferroni correction, a highly stringent statistical criterion ($P < 0.0005$), was applied to the dataset (Broadhurst and Kell 2006). It is often criticized as being too selective because meaningful biological data may be lost, however it reduces the possibility that the observed differences arose by chance. Only 7 metabolites varied between genotypes using this test (Table 2) and, when those compounds that were different in content due to over-expression of *Dx5* were considered, the list was further reduced to 3 metabolites (Table 2). They were guanine, 4-hydroxycinnamic acid and the uncharacterized compound, Unknown 071306a. A role for these compounds in primary metabolism in the endosperm is unclear. For example, 4-hydroxycinnamic acid is associated with cell wall metabolism (Caspi et al. 2006) and guanine reportedly accumulates to high levels in the embryo (Cheung and Marcus 1976). Because the entire caryopsis was used it is difficult to determine the extent to which these genotype-based differences are due to metabolic activity in embryo and pericarp as opposed to that in the endosperm where the *HMW-GS Dx5* is normally synthesized and stored. The other affected compound, Unknown_071306a, was not chemically identified and so we are unable to assign a role for it in caryopsis metabolism.

It is possible that changes introduced into the transgenic lines affected starch levels. Starch biosynthesis is the main metabolic activity in the developing wheat caryopsis (Stamova 2007) and changes in protein biosynthesis can

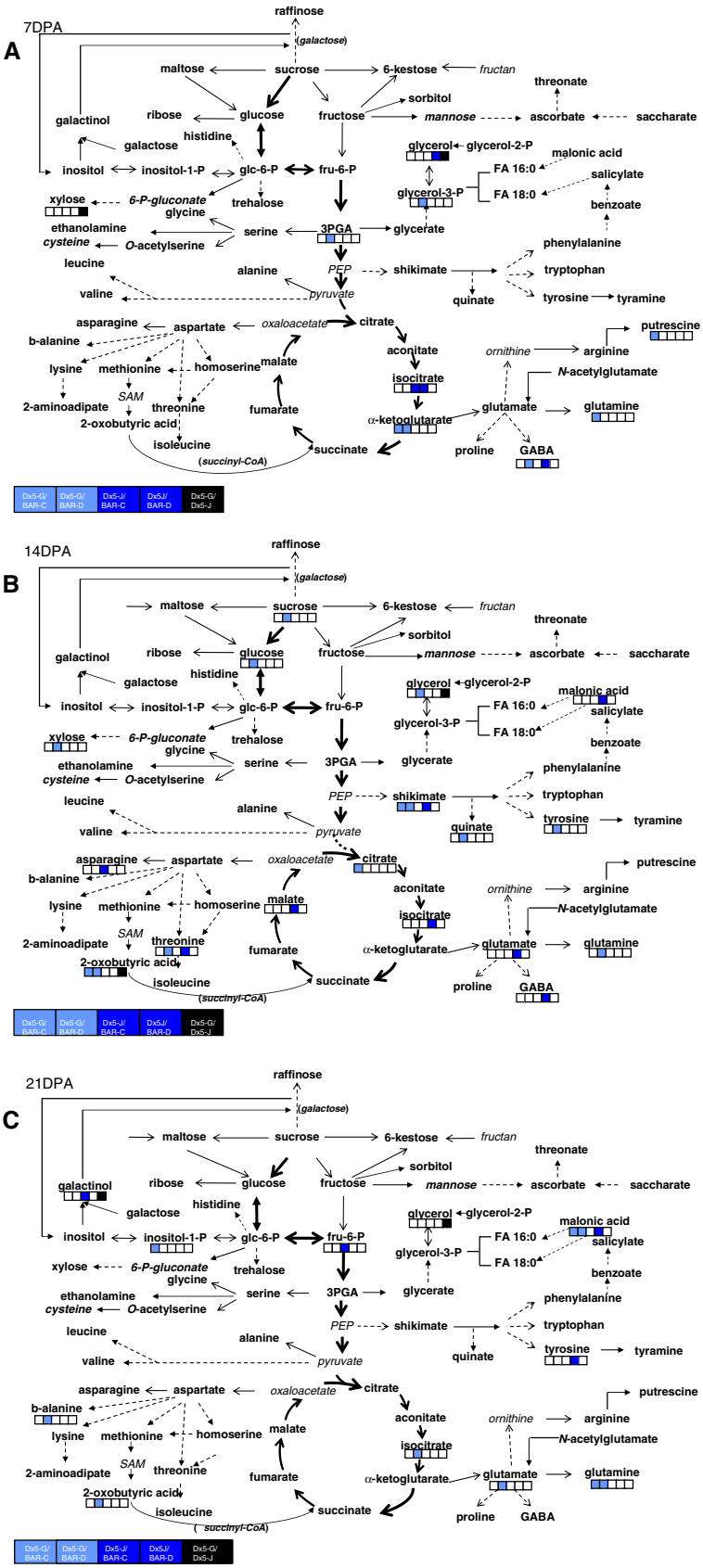


Fig. 4 Mapping metabolite changes between *Dx5*-transformed and *bar*-only genotypes. Each map represents a different developmental stage: **a** 7 DPA, **b** 14 DPA and **c** 21 DPA. Metabolites that were different ($P < 0.05$) between *Dx5*-transformed genotypes are shown on the map by coloured boxes. Metabolites that differed between Dx5-G and the two *bar*-only transformants: light squares; Dx5-J and the two *bar*-only transformants: dark squares; and metabolites that were different between Dx5-G and Dx5-J: black squares. Metabolites that differed between the two *bar*-only transformants are not shown. Metabolites that varied between the genotypes that could not be placed on the map are not shown and are in the Supplemental Fig. 2. There were 11, 16 and 10 differing compounds at 7, 14 and 21 DPA, respectively

Table 2 Metabolites with significantly different levels between genotypes using the parametric Student's *t*-tests

	BAR-C/BAR-D		Dx5-G/BAR-C		Dx5-J/BAR-C		Dx5-G/BAR-D		Dx5-J/BAR-D		Dx5-G/Dx5-J	
	av. ratio	*t*-test	av. ratio	*t*-test	av. ratio	*t*-test	av. ratio	*t*-test	av. ratio	*t*-test	av. ratio	*t*-test
7 DPA												
4-Hydroxycinnamic acid	1.57	**0.01065**	1.97	0.05954	1.34	0.13344	1.97	*0.00042*	1.34	**0.02210**	1.47	**0.00146**
Acetohydroxyamic acid	3.33	*0.00009*	2.43	**0.01303**	2.43	**0.00777**	2.43	**0.00047**	2.43	*0.00020*	1.00	0.98158
Guanine	1.66	**0.00673**	1.84	0.25004	0.99	**0.00115**	1.84	**0.00064**	0.99	0.92807	1.87	*0.00003*
Lactic acid	1.84	**0.00247**	3.23	**0.00117**	1.83	0.95903	3.23	*0.00002*	1.83	**0.00066**	1.76	**0.00061**
Unknown_071306a	2.04	**0.00237**	3.63	**0.00333**	2.78	0.12828	3.63	*0.00016*	2.78	**0.00392**	1.31	0.15340
14 DPA												
4-Hydroxycinnamic acid	1.54	0.08533	1.02	0.87540	0.83	0.24020	1.57	**0.04912**	1.29	0.26106	1.23	0.11510
Acetohydroxyamic acid	1.88	**0.02213**	0.55	**0.03556**	0.94	0.70199	1.04	0.87162		**0.00296**	0.59	**0.01017**
Lactic acid	1.67	**0.01681**	1.25	0.20008	1.17	0.28715	2.09	**0.00175**	1.96	*0.00040*	1.07	0.60958
U_021706b_33.7	1.55	0.08528	0.96	0.81593	1.23	0.56858	1.49	0.07455	1.91	0.15302	0.78	0.48331
Unknown_071306a	2.27	0.05565	1.68	0.19908	1.57	0.09360	3.83	**0.01836**	3.57	*0.00014*	1.07	0.80245
21 DPA												
1,6-Anhydroglucose	1.96	*0.00046*	0.82	0.42980	0.71	**0.01079**	1.61	0.20554	1.40	0.07949	1.15	0.64199
4-Hydroxycinnamic acid	1.54	**0.00122**	1.25	0.05887	0.87	0.29048	1.94	**0.00087**	1.34	0.10429	1.45	**0.03272**
Acetohydroxyamic acid	2.36	*0.00054*	0.78	0.14475	0.81	0.10593	1.84	**0.03467**	1.92	**0.00648**	0.96	0.81007
Heptadecanoic acid	0.99	0.94656	1.27	0.14915	0.90	0.53550	1.25	0.16027	0.89	0.48292	1.40	**0.04189**
Lactic acid	1.71	*0.00002*	1.78	*0.00038*	1.14	0.28333	3.04	*0.00002*	1.95	**0.00206**	1.56	**0.00673**
U_021706b_33.7	2.32	*0.00017*	0.70	**0.02779**	0.45	**0.00415**	1.62	**0.01276**	1.05	0.86413	1.55	0.10716
Unknown_071306a	2.29	**0.00179**	3.19	**0.00532**	1.34	0.18764	7.30	**0.00140**	3.06	**0.00443**	2.39	**0.01607**

This table represents a subset of the metabolites that accumulated to different levels among genotypes. The average ratio (av. ratio) was calculated as described in Sect. 2. Significant *t*-test values ($P < 0.05$) are marked in black and bold. All of the values corrected for false discovery (Bonferroni corrections $P < 0.0005$), are shown in this table and are marked in italics and bold

impact starch accumulation (Giroux et al. 1994). There was increased starch in both *Dx5* genotypes compared to the controls at 14 DPA (Fig. 5), but not at any other stage. The ratios of 3-phosphoglyceric acid (3-PGA) to phosphate, which are indicative of conditions that regulate starch biosynthesis (Tetlow et al. 2004), did not explain starch content variation between genotypes (data not shown). Differences in starch accumulation were probably a result of a change in some unknown regulatory mechanism in the developing caryopsis.

To sum up, the changes in metabolite levels brought on by *Dx5*-over-expression affected many pathways, but these were small. At $P < 0.0005$, only 3 metabolites were altered between the *Dx5*- and the control lines. Parametric tests and multivariate analyses all indicate that disparities in metabolite levels between genotypes were minimal (Fig. 1). Differences in metabolite and starch levels between the controls and *Dx5* genotypes also changed during development, with the most pronounced alterations were found at 14 DPA and the fewest at 7 DPA (Fig. 4).

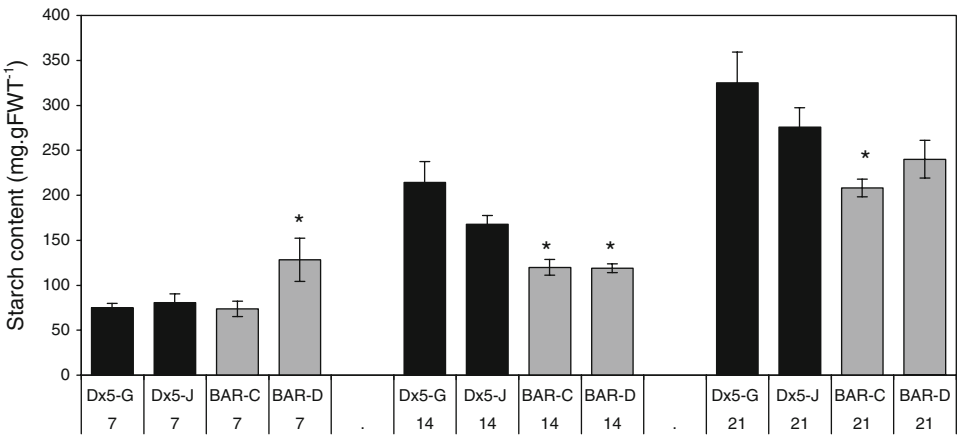


Fig. 5 Starch content of developing transgenic wheat caryopses. Starch content was measured using HPLC. Each bar is the mean ± standard error of the mean of 10 caryopses from 5 plants. Caryopses were sampled from the same plants used for GC–MS profiling. On the *x*-axis: first row-developmental stage in DPA; second row-plant genotype. An asterisk indicates statistically significant differences between genotypes

Gluten deposition begins at 14DPA in wild-type wheat endosperm (Pomeranz 1988) and hence few metabolic effects of over-expression of *Dx5* might be detected before then, i.e. at 7DPA. It is possible that in these transgenic lines most of the mechanistic adjustments to Dx5 levels occurred around 14 DPA, including an effect on starch biosynthesis.

3.4 Differences in the structure of the polar metabolic networks between genotypes

The structure of a metabolic network can be assessed by examining its metabolite-to-metabolite ratios. These ratios are the result of the stoichiometric relationships of several biochemical reactions which are robust (ap Rees and Hill 1994; Morgenthal et al. 2006); only a fundamental disturbance of the metabolic network would disrupt these established correlations (Fell 1997; Steuer et al. 2003; Steuer 2006). It therefore follows that changes in metabolite ratios may be better indicators of biochemical pathway perturbations than fold-changes in metabolite levels between samples (Steuer et al. 2003; Morgenthal et al. 2006; Steuer 2006). Metabolite correlative analysis was effectively used by Weckwerth et al. (2004) to identify changes in metabolism in an *Arabidopsis* mutant that had a "silent phenotype", i.e. showed no obvious metabolic or phenotypic effect of the mutation. They were able to detect carbohydrates and amino acids with altered profiles that differed in the mutant and which may explain the result of the perturbation.

To determine potential structural differences in metabolic networks due to over-expression of *Dx5* we performed a similar analysis as that done by Weckwerth et al. (2004). Pair-wise comparisons of the 109 metabolites to each other were calculated using Pearson's correlation coefficients. The resulting matrix is called the metabolite-to-metabolite correlation matrix. Heat maps drawn using these matrices give a bird's-eye view of the unique signature of the metabolic networks that define each genotype (Fig. 6a–d). From the data generated, several observations can be made. As expected, most (76–86%) of the metabolic correlative patterns were positive in each genotype, i.e. $r > 0$, however BAR-D had more positive correlations than the others (Table 3). Metabolites with strong positive correlations ($r > 0.95$) have a high probability of being connected biologically and may be under similar regulatory control (Weckwerth and Fiehn 2002; Steuer 2006, 2007). These metabolites represented 0.72–1.22% of the total possible (5886) and, noticeably, fewer were found in Dx5-J (Table 3). However at a lower *r*-value, i.e. >0.9, the number of positive correlations was comparatively higher in Dx5-J than the other genotypes (Table 3). The higher variance found in Dx5-J could reduce the number of the strongest correlations and partially explain this result. Still, as pointed out previously, variance was similar between the two Dx5-genotypes, yet Dx5-G behaves more similarly to the controls, and Dx-5J differs, as in this analysis.

Another way of assessing the extent to which the genotypes differ is to identify the number of metabolite that did not correlate with others. As shown in Table 4, the greatest contrast was between Dx5-J and the controls (Table 4). A corollary of this analysis is to identify metabolites that were negatively correlated, as a strong negative correlation ($r > -0.8$) between metabolites may be due to different regulatory mechanisms acting at those metabolic steps when compared to others (Weckwerth and Fiehn 2002). These are shown in Table 5, there were three negatively correlated pairs in Dx5-G and none in BAR-C. There were no overlapping compounds between genotypes and the role some of them play in metabolism (maleic acid and 2,3-dihydrobutanedioic acid) is not clear. Galactinol was negatively correlated with shikimic acid in Dx5-J and was also a determinant of the PCA separation (Fig. 2 and Supplemental Table 1), however the significance of this observation is not obvious. We also identified metabolites that are positively correlated in the controls but negatively correlated in the *Dx5* genotypes (or vice-versa) as they may indicate changes that result from an abrupt switch or reversal in a regulatory mechanism due to over-expression of *Dx5*. Glycerol-2-P and tetracosanoic acid were negatively correlated in Dx5-G ($r = -0.329$) and Dx5-J ($r = -0.340$) but were positive in BAR-C ($r = 0.520$) and BAR-D ($r = 0.477$). This was also true of threonine and UL028 where the Pearson's *r*-values were −0.198, −0.230, 0.552 and 0.478 among Dx5-G, Dx5-J, BAR-C and BAR-D, respectively. It is not possible to determine how, or if these compounds are related biochemically, but they represent a change in the genotypes that appears to be dependent on *Dx5*-expression.

If the correlative analyses are an accurate reflection of the metabolite networks in the samples studied then some differences were identified between genotypes. The metabolite network of the Dx5-J genotype appeared to be most perturbed having the least number of strongly correlating metabolites and the most differences when compared to the controls. The caveat is that the variance in the sample was highest for this genotype, however, it was only slightly higher than Dx5-G which did not vary substantially from the controls by correlative analysis. It is possible therefore that over-expression of Dx5 contributed to the altered metabolite network in Dx5-J.

3.5 Metabolites with altered developmental profile between genotypes

Having looked at the overall topology of the network in the genotypes, we then focused on identifying the specific

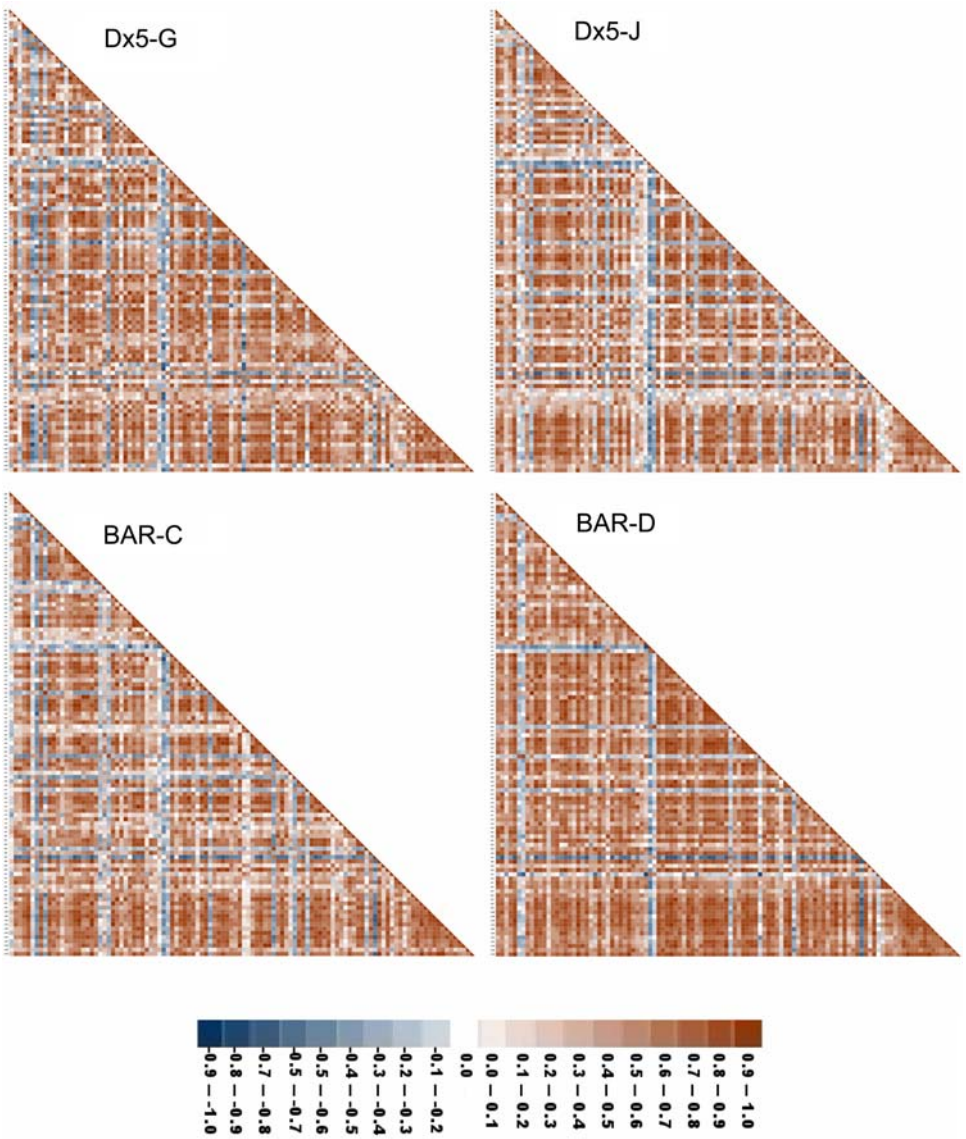


Fig. 6 Heat-map signatures of the metabolite correlative matrices for each genotype. Heat maps for each genotype were generated by computing the Pearson correlation coefficient for each metabolite-to-metabolite comparison. Each square on the heat map represents the correlation coefficient or linear relationship that results when two metabolites are compared to each other in the 5 biological replicates at the three developmental stages (15 samples total). The number of metabolite-to-metabolite pairs analysed for the correlative matrices of each genotype was calculated as follows $[(109*109)/2 - (109/2)] = 5886$. The heat maps for each genotype are as follows: **a** Dx5-G; **b** Dx5-J; **c** BAR-C; **d** BAR-D. Metabolites that did not correlate with others are shown in white, positive and negative correlations are shown as darker hues

metabolites that were altered. Such metabolites may represent points where the metabolic networks have broken down and thus help in understanding how metabolism was disrupted. There were a total of 47 such metabolites derived from the correlative analysis (Supplemental Table 3). Metabolites with modulated profiles in the *Dx5*-transformed lines most likely arose from disturbances due to *Dx5* expression. The ratios of proline, tetracosanoic acid and glyceric acid to other metabolites were different in each of the *Dx5*-transformed genotypes when compared to either of the control lines or to each other. Shikimic acid and threonic acid-1,4-lactone levels were changed in all comparisons except Dx5-G to Dx5-J, which may indicate that larger differences in *Dx5* levels would be needed to show an analogously altered profile between these lines. *N*-acetylglutamate, a putative erythronic acid, and γ-aminobutyric acid (GABA) differed in 5 of the 6 genotype-to-genotype comparisons. There was no immediately obvious connection between these metabolites, however, upon careful examination, half of them: GABA, proline, glyceric acid and *N*-acetylglutamine, are metabolically linked to glutamine and glutamine (Fig. 3a–c). These amino acids are the entry point for the biosynthesis of other amino acids and, more importantly, make up 45% of the dry matter of the Dx5 and other glutenin proteins (Pomeranz 1988). Detailed studies of the storage protein composition of the Dx5-G and Dx5-J lines indicated that there was an increase in the percentage of total polymeric proteins which are glutamine- and glutamate-rich (Blechl et al. 2007). It is intriguing to speculate that the pathways immediately connected to these "core amino acids" were adjusted to meet the changes in their soluble pools when used for Dx5 synthesis.

Table 3 Metabolite-to-metabolite relationships within each genotype

Correlations	Dx5-G	Dx5-J	BAR-C	BAR-D
$0.95 \leq r \leq 1.0$	73	43	69	70
$0.90 \leq r < 0.95$	195	212	165	263
$0.80 \leq r < 0.90$	478	499	401	644

The number of positively correlating metabolites within each genotype. Correlations were calculated using Pearson rank correlation coefficient; the *r*-value shows the strength of the relationships. A total of 5886 correlations were possible

3.6 Metabolite profiles of unknown polar compounds

Some compounds that have not been structurally defined showed significant differences between genotypes (Fig. 4 and Table 4), and one of them (UL028) was identified by DFA as determining the separation between genotypes (Fig. 3a). It would be interesting to determine a potential role for these compounds in central metabolism by correlative analysis (Fig. 7). The strongest correlations for these unknown compounds were found with metabolites involved in carbohydrate and fatty acid metabolism (7a–c). Examination of their spectral features show that U_021706b_33.7 and Unknown_071306a are not related to sugars or fatty acids. The strong correlations we observed may be the result of connections, through a common precursor or regulatory mechanism, to compounds of central metabolism. Determining the structures of these "unknown" compounds may help to elucidate their roles, if any, in biological processes in wheat caryopses.

3.7 Effect of gene transformation on metabolic networks

A secondary aim of this work was to assess the extent to which genetic changes due to variations produced by the gene transformation process itself can alter metabolism in wheat caryopses. Genetic manipulations through in vitro techniques can introduce spurious but heritable variability. The expression of the selectable marker genes, the place(s) where the plasmid integrates in the genome as well as plasmid architecture and copy number and epigenetic changes due to somaclonal variation, may all have unpredictable effects on the transgenic organism (Barcelo et al. 2001). Several results point to significant genetic changes that were not attributable to *Dx5* expression but could be due to variation arising from gene transformation. First, the number of metabolites that were altered when BAR-C and BAR-D were compared was in the same range seen as when they were compared to the *Dx5*-transformed genotypes (data not shown). Second, several compounds that varied in content between *Dx5*-transformed genotypes and the *bar*-only controls also differed when the two controls were compared to each other (Table 4 and Supplemental Table 4). Seven compounds differed between the genotypes after correcting for false discovery (Table 2), and over half (four) were also altered when the two *bar*-only controls were compared side-by-side. Third, of the 47 metabolites with altered profiles identified from the correlative analysis, 22 varied when BAR-C and BAR-D were compared to each other and therefore may be due to

Table 4 Number of metabolites that do not correlate ($r < 0.90$) when compared between genotypes

Genotypes compared $r < 0.90$	Dx5-G vs. BAR-C	Dx5-G vs. BAR-D	Dx5-G vs. Dx5-J	Dx5-J vs. BAR-C	Dx5-J vs. BAR-D	BAR-C vs. BAR-D
Number of metabolites	21	23	29	32	31	23

The response ratios (the average metabolite levels) of each of the 109 metabolites were compared for each combination of pairs of genotypes and at all developmental stages. Pearson rank correlation coefficient was used to determine the extent to which they correlated and the number that did not for each pair are indicated here

Table 5 Negatively correlating metabolite pairs in the wheat transgenic genotypes

Genotype	Negatively correlating metabolite pair		*r*-value	*P*-value
BAR-C	–	–		
BAR-D	t-Ferulic acid	Maleic acid	−0.823	1.6E−04
Dx5-G	2,3-Dihydroxybutanediol	Tryptophan	−0.804	3.0E−04
	2,3-Dihydroxybutanediol	UL028	−0.844	7.6E−05
	1-Inositol phosphate	Putrescine	−0.811	2.4E−04
Dx5-J	Shikimic acid	Galactinol	−0.815	2.1E−04

Metabolites that were negatively associated in each wheat genotype as determined by Pearson's rank correlations. An all-to-all comparison of the 109 metabolites was done and those showing a strong negative correlation are shown. Each correlations has a *P*-value below 0.05/109 = 4.6E−04 (Bonferroni correction) and are statistically significant

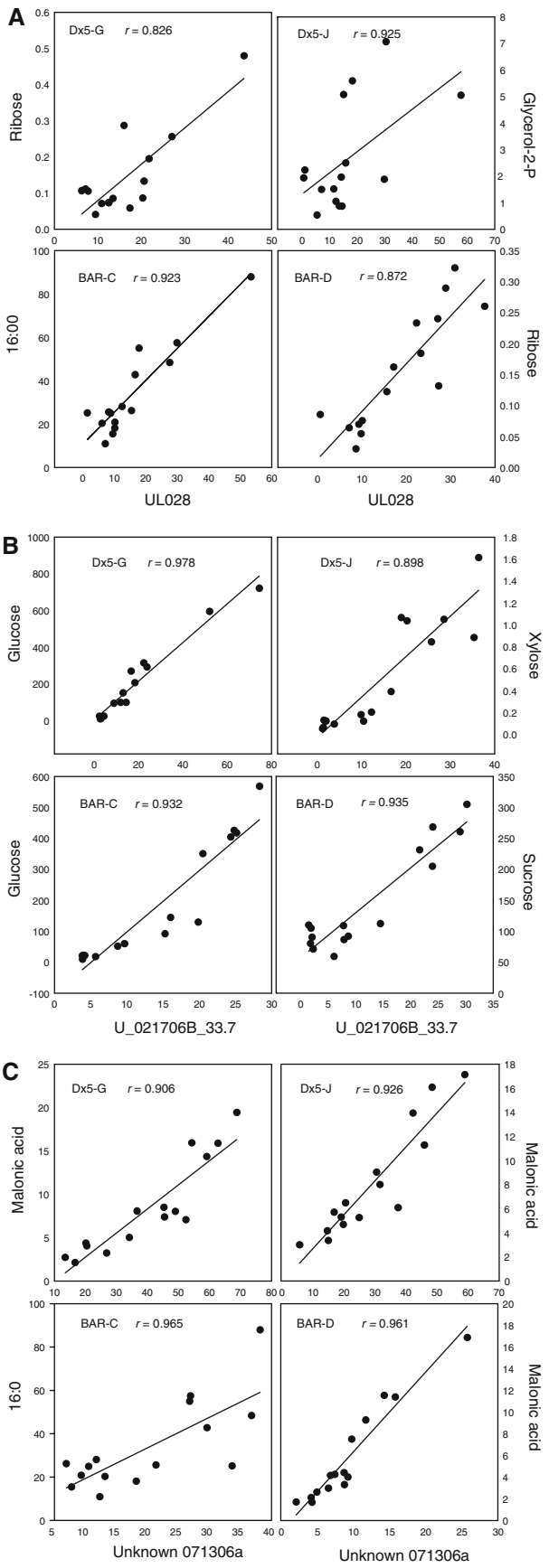


Fig. 7 Correlative scatter plots of metabolite levels in transgenic ▶ wheat genotypes. Each graph depicts a single compound with the highest correlative pattern when compared with one of three unknown compounds. Graphs (**a**–**c**) show the correlation of **a** UL028; **b** U_021706b_33.7 and **c** Unknown_071306a with other metabolites in each genotype. There are 15 points on each graph and an individual point or symbol represents metabolite level measured in one of the five biological replicates at one of the three developmental stages studied. The levels of the unknown compounds were identified as altered due to over-expression of *Dx5* in different statistical tests. The strength of the correlation is indicated by the regression value (*r*) on each graph. The *P*-value for each correlation was <1.0E−4 and are likely to be statistically significant

genetic changes associated with the transformation process (Supplemental Table 4). These compounds included carbohydrates (5), amino acids (5), TCA intermediates (3), fatty acids or fatty acid intermediates (3), and compounds not involved in primary metabolic pathways (6), illustrating the broad and non-specific nature of the metabolic changes induced by transformation. We are unable to definitively pinpoint the underlying cause of differences between the selectable-marker controls, as this was beyond the scope of our study. Even so, these results underscore the findings in the literature that the effects of gene transformation, i.e. position effects, level of selectable-marker expressed and/or somaclonal variation, can influence physiological and biochemical processes in transgenic organisms (Filipecki and Malepszy 2006). The fact that over-expression of *Dx5* showed a "silent phenotype" perhaps made identifying the effect of gene transformation easier. Still, this conclusion is based only on the measurements of a small number of highly abundant metabolites in the caryopsis and may not be representative of broader changes to the metabolome.

4 Conclusion

Given the plasticity and complexity of plant metabolism it is not possible to make sweeping conclusions about the effects of changing Dx5 expression in wheat caryopses from a single study. Therefore this work complements and extends existing knowledge of how over-expression of *Dx5* affects wheat metabolism. Baker et al. (2006) comprehensively assessed, using metabolite fingerprinting and GC–MS, the effect of the environment on *Dx5*-transformed genotypes. They used flour milled from ripened caryopses bulked from several plants. In our study two independently derived transformants over-expressing *Dx5* and two independently derived *bar*-only transformed genotypes were used, to provide insights on the potential effect of gene transformation as well as those of the specified perturbation. Approximately 50% of the observed changes in metabolites were caused by genetic changes from plant

transformation. Furthermore, metabolites in caryopses from five individual plants were assayed to observe plant-to-plant variation, which was as high as the variation between genotypes. Analysis of tissues of different developmental stages suggested that changes in metabolite levels "normalised" with maturation. Our multivariate analysis showed that in spite of the similarity between genotypes by PCA and parametric tests, the genotype with the highest level of Dx5 could be differentiated from the others.

Acknowledgements We thank Dr. Ann Blechl for the generous gift of the transgenic seeds and information on the lines. We are indebted to Dr. Ron Haff, USDA-ARS Albany, for help with SAS Software and to Nick Petkov and Dobromir Tzankov for technical assistance. We thank Drs. Belinda Martineau and Olin Anderson for comments on the manuscript. This work was supported by grants to the Australian Centre for Plant Functional Genomics from the Grains Research and Development Corporation, the Australian Research Council, the South Australian Government, the University of Adelaide and the University of Melbourne (SS, UR, AB); and to Metabolomics Australia provided by the Australian Government through the National Collaborative Research Infrastructure Strategy (UR, AB), USDA-ARS CRIS Project 5325-21000-011 (DLC, BS) and National Science Foundation Grant: NSF-MCB-0620001 (DMB).

References

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 1–9). Thousand Oaks, CA: Sage.

Altpeter, F., Popelka, J. C., & Wieser, H. (2004). Stable expression of 1Dx5 and 1Dy10 high-molecular-weight glutenin subunit genes in transgenic rye drastically increases the polymeric glutelin fraction in rye flour. *Plant Molecular Biology, 54*, 783–792. doi:10.1007/s11103-004-0122-5.

Alvarez, M. L., Guelman, S., Halford, N. G., Lustig, S., Reggiardo, M. I., Ryabushkina, N., et al. (2000). Silencing of HMW glutenins in transgenic wheat expressing extra HMW subunits. *Theoretical and Applied Genetics, 100*, 319–327. doi:10.1007/s001220050042.

ap Rees, T., & Hill, S. A. (1994). Metabolic control analysis of plant-metabolism. *Plant, Cell & Environment, 17*, 587–599. doi:10.1111/j.1365-3040.1994.tb00151.x.

Baker, J. M., Hawkins, N. D., Ward, J. L., Lovegrove, A., Napier, J. A., Shewry, P. R., et al. (2006). A metabolomic study of substantial equivalence of field-grown genetically modified wheat. *Plant Biotechnology Journal, 4*, 381–392. doi:10.1111/j.1467-7652.2006.00197.x.

Barcelo, P., Rasco-Gaunt, S., Thorpe, C., & Lazzeri, P. A. (2001). Transformation and gene expression. *Advances in Botanical Research Incorporating Advances in Plant Pathology, 34*(34), 59–126.

Barro, F., Barcelo, P., Lazzeri, P. A., Shewry, P. R., Martin, A., & Ballesteros, J. (2002). Field evaluation and agronomic performance of transgenic wheat. *Theoretical and Applied Genetics, 105*, 980–984. doi:10.1007/s00122-002-0996-z.

Baudo, M. M., Lyons, R., Powers, S., Pastori, G. M., Edwards, K. J., Holdsworth, M. J., et al. (2006). Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. *Plant Biotechnology Journal, 4*, 369–380. doi:10.1111/j.1467-7652.2006.00193.x.

Beckles, D. M., Smith, A. M., & ap Rees, T. (2001). A cytosolic ADP-glucose pyrophosphorylase is a feature of *Graminaceous* endosperms, but not of other starch-storing organs. *Plant Physiology, 125*, 818–827. doi:10.1104/pp.125.2.818.

Blechl, A., Lin, J., Nguyen, S., Chan, R., Anderson, O. D., & Dupont, F. M. (2007). Transgenic wheats with elevated levels of Dx5 and/or Dy10 high-molecular-weight glutenin subunits yield doughs with increased mixing strength and tolerance. *Journal of Cereal Science, 45*, 172–183. doi:10.1016/j.jcs.2006.07.009.

Blechl, A. E., & Anderson, O. D. (1996). Expression of a novel high-molecular-weight glutenin subunit gene in transgenic wheat. *Nature Biotechnology, 14*, 875–879. doi:10.1038/nbt0796-875.

Blechl, A. E., Le, H. Q., & Anderson, O. D. (1998). Engineering changes in wheat flour by genetic transformation. *Journal of Plant Physiology, 152*, 703–707.

Bregitzer, P., Blechl, A. E., Fiedler, D., Lin, J., Sebesta, P., De Soto, J. F., et al. (2006). Changes in high molecular weight glutenin subunit composition can be genetically engineered without affecting wheat agronomic performance. *Crop Science, 46*, 1553–1563. doi:10.2135/cropsci2005.10-0361.

Broadhurst, D. I., & Kell, D. B. (2006). Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics, 2*, 171–196. doi:10.1007/s11306-006-0037-z.

Caspi, R., Foerster, H., Fulcher, C., Hopkinson, R., Ingraham, J., Kaipa, P., et al. (2006). MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research, 34*, D511–D514. doi:10.1093/nar/gkj128.

Cheung, C. P., & Marcus, A. (1976). Guanine nucleotide determination in extracts of wheat embryo. *FEBS Letters, 70*, 141–144. doi:10.1016/0014-5793(76)80744-2.

Christensen, A. H., & Quail, P. H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research, 5*, 213–218. doi:10.1007/BF01969712.

FAO. (2006). Wheat. *FAO statistical yearbook*.

Fell, D. (1997). *Understanding the control of metabolism*. London: Portland Press.

Filipecki, M., & Malepszy, S. (2006). Unintended consequences of plant transformation: A molecular insight. *Journal of Applied Genetics, 47*, 277–286.

Fu, D. L., Uauy, C., Blechl, A., & Dubcovsky, J. (2007). RNA interference for wheat functional gene analysis. *Transgenic Research, 16*, 689–701. doi:10.1007/s11248-007-9150-7.

Giroux, M. J., Boyer, C., Feix, G., & Hannah, L. C. (1994). Coordinated transcriptional regulation of storage product genes in the maize endosperm. *Plant Physiology, 106*, 713–722.

He, G. Y., Rooke, L., Steele, S., Bekes, F., Gras, P., Tatham, A. S., et al. (1999). Transformation of pasta wheat (*Triticum turgidum* L-var. durum) with high-molecular-weight glutenin subunit genes and modification of dough functionality. *Molecular Breeding, 5*, 377–386. doi:10.1023/A:1009681321708.

Jacobs, A., Lunde, C., Bacic, A., Tester, M., & Roessner, U. (2007). The impact of constitutive heterologous expression of a moss Na+ transporter on the metabolomes of rice and barley. *Metabolomics, 3*, 307–317. doi:10.1007/s11306-007-0056-4.

Joachimiak, M. P. J., Weissman, J. L., & May, B. C. H. (2006). JColorGrid: Software for the visualizaton of biological measurements. *BMC Bioinformatics, 7*, 225. doi:10.1186/1471-2105-7-225.

Kermit, M., & Tomic, O. (2003). Independent component analysis applied on gas sensor array measurement data. *IEEE Sensors Journal, 3*, 218–228. doi:10.1109/JSEN.2002.807488.

Laudencia-Chingcuanco, D. L., Stamova, B. S., You, F. M., Lazo, G. R., Beckles, D. M., & Anderson, O. D. (2007). Transcriptional profiling of wheat caryopsis development using cDNA microarrays. *Plant Molecular Biology, 63*, 651–668. doi:10.1007/s11103-006-9114-y.

Morgenthal, K., Weckwerth, W., & Steuer, R. (2006). Metabolomic networks in plants: Transitions from pattern recognition to biological interpretation. *Bio Systems, 83*, 108–117. doi:10.1016/j.biosystems.2005.05.017.

Payne, P. I., Nightingale, M. A., Krattiger, A. F., & Holt, L. M. (1987). The relationship between Hmw glutenin subunit composition and the bread-making quality of british-grown wheat-varieties. *Journal of the Science of Food and Agriculture, 40*, 51–65. doi:10.1002/jsfa.2740400108.

Pomeranz, Y. (1988). *Wheat: chemistry and technology*. St. Paul: American Association of Cereal Chemists.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N. S., Hayes, A., Walsh, M. C., et al. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology, 19*, 45–50. doi:10.1038/83496.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., et al. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *The Plant Cell, 13*, 11–29.

Roessner, U., Patterson, J. H., Forbes, M. G., Fincher, G. B., Langridge, P., & Bacic, A. (2006). An investigation of boron toxicity in barley using metabolomics. *Plant Physiology, 142*, 1087–1101. doi:10.1104/pp.106.084053.

Rooke, L., Steele, S. H., Barcelo, P., Shewry, P. R., & Lazzeri, P. A. (2003). Transgene inheritance, segregation and expression in bread wheat. *Euphytica, 129*, 301–309. doi:10.1023/A:1022296017801.

Scholz, M., Gatzek, S., Sterling, A., Fiehn, O., & Selbig, J. (2004). Metabolite fingerprinting: Detecting biological features by independent component analysis. *Bioinformatics (Oxford, England), 20*, 2447–2454. doi:10.1093/bioinformatics/bth270.

Shewry, P. R., Gilbert, S. M., Savage, A. W. J., Tatham, A. S., Wan, Y. F., Belton, P. S., et al. (2003). Sequence and properties of HMW subunit 1Bx20 from pasta wheat (Triticum durum) which is associated with poor end use properties. *Theoretical and Applied Genetics, 106*, 744–750.

Shewry, P. R., Halford, N. G., Belton, P. S., & Tatham, A. S. (2002). The structure and properties of gluten: An elastic protein from wheat grain. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 357*, 133–142.

Stamova, B. S. (2007). *Gene-to-metabolite networks in developing wheat caryopsis*. Ph.D. Thesis, University of California-Davis, p. 150.

Statsoft. (2003). *Statistica dataminer*.

Steel, R. G. D., Torrie, J. H., & Dickey, D. A. (1997). *Principles and procedures of statistics a biometrical approach*. New York: McGraw-Hill.

Steuer, R. (2006). On the analysis and interpretation of correlations in metabolomic data. *Briefings in Bioinformatics, 7*, 151–158. doi:10.1093/bib/bbl009.

Steuer, R. (2007). Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry, 68*, 2139–2151. doi:10.1016/j.phytochem.2007.04.041.

Steuer, R., Kurth, J., Fiehn, O., & Weckwerth, W. (2003). Observing and interpreting correlations in metabolomic networks. *Bioinformatics (Oxford, England), 19*, 1019–1026. doi:10.1093/bioinformatics/btg120.

Sweetlove, L. J., & Fernie, A. R. (2005). Regulation of metabolic networks: Understanding metabolic complexity in the systems biology era. *The New Phytologist, 168*, 9–23. doi:10.1111/j.1469-8137.2005.01513.x.

Tetlow, I. J., Morell, M. K., & Emes, M. J. (2004). Recent developments in understanding the regulation of starch metabolism in higher plants. *Journal of Experimental Botany, 55*, 2131–2145. doi:10.1093/jxb/erh248.

Vain, P., James, V. A., Worland, B., & Snape, J. W. (2002). Transgene behaviour across two generations in a large random population of transgenic rice plants produced by particle bombardment. *Theoretical and Applied Genetics, 105*, 878–889. doi:10.1007/s00122-002-1039-5.

Vasil, I. K. (2007). Molecular genetic improvement of cereals: Transgenic wheat (*Triticum aestivum* L.). *Plant Cell Reports, 26*, 1133–1154. doi:10.1007/s00299-007-0338-3.

Weckwerth, W., & Fiehn, O. (2002). Can we discover novel pathways using metabolomic analysis? *Current Opinion in Biotechnology, 13*, 156–160. doi:10.1016/S0958-1669(02)00299-9.

Weckwerth, W., Loureiro, M., Wenzel, K., & Fiehn, O. (2004). Differential metabolic networks unravel the effects of silent plant phenotypes. *Proceedings National Academy Science USA, 101*, 7809–7814.